University of Lincoln

College of Science

School of Life Sciences

Biochemistry and Molecular Biology

MSc by Research

Gene Co-expression of Mitosis Related Genes is a Stronger Indicator of Functional Engagement than Average Expression Level: A Comparison Involving High and Low Proliferating Tissues.

February 2015

By Beau Reynolds
REY10173147

Name of supervisor:
Dr. Humberto Gutierrez

A Postgraduate Research Project Thesis Submitted for Fulfilment of the Award of Masters of Science Degree at the University of Lincoln.

UNIVERSITY OF
LINCOLN

This work is dedicated to a dear friend, James “Algorn” Ryan, who passed away October 2013. Thank you for all your help with my first steps into the programming world, you will be sorely missed.

List of Abbreviations:

Mos - Months

Yrs – Years

MT – Mitosis

BG – Background

Pre – Pre-natal

Post – Post-natal

Ctrl – Control

NT – Nervous Tissues

NNT – Non-Nervous Tissues

CL – Cell Lines

TOM – Topological Overlap Matrix

GO – Gene Ontology

S.E.M – Standard Error of the Mean

Seq - Sequencing

Abstract:

Genes play a major role in controlling a wide range of biological, physiological and cellular processes. While the regulated level of expression of individual genes is known to play a key role in modulating specific cellular functions, genes however do not act in isolation. Instead, any individual cellular function is the result of a large number of genes acting in close coordination. In this study, we investigate the link between the relative level of gene expression or, alternatively, the level of co-ordination of functionally related genes (co-expression) and the level of demand for their associated biological function. Looking at mitosis genes, as defined by their Gene Ontology categorization, as a reference set of genes, we examined their expression and co-expression across tissue samples known to differ in their cell proliferation demands. Specifically, we looked at expression data from a variety of human-derived tissues including brain-derived samples at several developmental stages. Using microarray data, we compared highly proliferative cell lines with the much less proliferative nervous tissues and found that there was a significant decrease for both average expression and co-expression, of mitosis-associated genes, in line with decreased demand of proliferative functions. Co-expression, however, was a much stronger indicator of functional engagement. The observed contrast is not the result of comparing tumoural tissues with normal tissues as proliferative non-nervous tissues, showed significantly higher co-expression of mitosis genes than in nervous tissues. Using both microarray and RNA sequencing expression data, we compared average expression and co-expression of mitosis genes across pre-natal and post-natal brain-derived samples. In pre-natal development neurogenesis is higher as this is required for early formation of the brain, after birth this function is much less required. Our results show that absolute level of expression did not significantly alter, however co-expression of mitosis-associated genes was once again significantly higher during pre-natal development compared to post-natal stages. Our results show that co-expression is a robust signature of mitotic function requirement with average level of expression representing a less consistent predictor of functional activation. Whether this association between functional engagement and co-expression can be extended to all biological functions should be further investigated.

Acknowledgements:

The author would like to thank Humberto Gutierrez for his opportunity to undertake a research degree and the continued and unwavering support given throughout the project.

Extended thanks are given to the various research teams of Brainspan, BioGPS and Gene Expression Omnibus for use of their gene expression databases, which without, this research would not have been possible.

A huge thank you is also given to Mr and Mrs Reynolds to whom provided funding to undertake this research degree course.

Table Of Contents

Introduction

Genes constitute the blueprint that ultimately results in the production of specific proteins responsible for all biological processes involved in the development, maintenance and growth of an organism. However, different phenotypic traits display different levels of complexity, ranging from protein binding affinities, to complex developmental and behavioural patterns in multicellular organisms, making it difficult to establish the precise relationship between the activation of individual genes and their phenotypic expression. Indeed, complex phenotypes, however, are never the result of individual proteins acting in isolation. Instead, they are the result of a wide array of complex biochemical processes directed by an equally large number of individual proteins and their associated genes. As a result, a complex interplay of numerous genes and proteins is needed in order to bring about most biological processes including formation of molecular complexes that influence the physical and chemical properties of numerous metabolites, altering or regulating the secondary and tertiary structure of other proteins, assembling complex signalling and regulatory circuits or directing the assembly of large molecular and anatomical structures at the organismal level (Tolia and Joshua-Tor 2006). These complex “cocktails” of proteins are required to ensure that biological functions are carefully assembled, activated and controlled to bring about normal cellular and physiological functions. How genes dynamically interact, at the level of expression, with each other to instruct complex functions is currently poorly understood.

In this study we set out to investigate the relationship between the functional engagement of a set of genes known to be involved in a common function and their collective pattern of expression.

Specifically we examined changes in the overall level of expression, on the one hand, or the overall level of coordinated activity (or co-expression) on the other, of mitosis-associated genes across tissues and developmental conditions differing in the level of mitotic activity. A study of this nature has never been carried out before.

While absolute level of expression has traditionally been used as an indicator of increased functional engagement of individual genes, whether the level of coordinated regulation, or co-expression, better reflects increased functional engagement of an assembly of genes has never been formally explored.

Coordinated gene expression can be readily measured by looking at existing correlations in expression levels between groups of genes across a series of suitable chosen tissue samples. Clustering analysis based on co-expression patterns has been used to identify groups or modules of correlated genes that may form molecular complexes, pathways or contribute in common regulatory and signalling circuits.

Gene co-expression analysis has been widely used to gain insights into the functional interactions between genes and the wider regulatory organization of transcriptomes across tissues, conditions and species. (Zhang and Horvath 2005, Oldham, Horvath et al. 2006, Oldham, Konopka et al. 2008, Saris, Horvath et al. 2009, Usadel, Obayashi et al. 2009, Torkamani, Dean et al. 2010, Obayashi and Kinoshita 2011). Apart from revealing functional networks of interacting genes, gene co-expression also provides information on the regulatory structure associated to a global expression profile as co-expressed genes are likely to be under the concerted control of a common complement of transcriptional regulators

It has been shown that pairs of genes displaying high co-expression values tend to share transcription factor (TF) binding sites. Most of these TF binding site sharing genes are also known to be associated to common functions (Marco, Konikoff et al. 2009). This suggests, but does not prove, that the activation of gene regulators involved in a given function should lead to an increased co-expression of the associated regulated genes. In this work we will be focusing on mitosis-related genes as a model gene set, given the well documented activation of mitotic functions in various tissues and developmental stages independently of the knowledge of the pattern of expression of associated genes. In this way we can directly assess the extent to which the level of expression and/or co-expression of these genes align with the level of activity of mitotic functions. Thus, by looking at and comparing gene expression data for mitosis-related genes across tissues with high or low proliferative activity we can test the potential association between the expression and co-expression of these genes and their actual functional engagement.

In our first study we will look at expression data from highly proliferative tumoural cell lines, and compare them against post-natal nervous tissues. Cell lines are known to be highly mitotically active whereas post-natal nervous tissue is known to be the least proliferative tissue in mammals (Tramontin, Garcia-Verdugo et al. 2003).

Following this, using tissues obtained from pre-natal and post-natal nervous samples a similar comparison of co-expression and expression was conducted. At this point, we took advantage of the differential level of cell proliferation activity across different developmental stages as post-natal brain tissue is much less proliferative than the developing (pre-natal) nervous system (Cooper 2000).

Our results show that mitosis genes are highly co-expressed during periods of high engagement of proliferative functions and that the average expression level of mitosis-related genes is less consistent with the relative level of proliferative activity across tissues or developmental stages; suggesting that it is a poor indicator of functional engagement. Whether the association between level of co-expression and functional engagement is restricted to mitosis-related genes or can be extended to all cellular and physiological functions should be further explored.

Methods

Gene Annotation and ID conversion

All databases were converted to Ensembl ID's (if not already) using Biomart. Probesets that shared the same Ensembl ID were averaged; probesets that had no corresponding Ensembl ID were removed and probesets that mapped to multiple Ensembl IDs were also removed.

Gene Expression sources, tissue selection and Database Fragmentation

Gene expression data sources are listed in table 1.

The databases were split into two equally sized sections for the two different states. Fourteen nervous tissues, non-nervous tissues and cell lines were selected from the BioGPS database (Microarray datasets 1a and b, see table 1 and table 2). All three immediate pre-natal time points were selected (ages 16, 21 and 24 post-conception weeks), and three post-natal time points were selected (ages 19, 21 and 23) for the Brainspan database (RNA-seq dataset 3, table 1). All eight pre-natal time points were used and the first eight post-natal time points (4 months – 22 years) were used for the Yale University database (Microarray dataset 2, table 1).

Gene ontology annotations for mitosis genes (GO:0007067) were obtained from the Gene Ontology project database (www.geneontology.org), and these genes were isolated and stored in a separate file. These genes will be referred to as "mitosis" and the remaining genes will be referred to "background" genes.

A separate set of databases were created for GEO and Brainspan databases (Table 1, datasets 2 and 3) in which data was split into two sets containing the expression data of each tissue sample for pre-natal and post-natal time windows

respectively. This set of databases was also used to examine time course of expression and co-expression.

Gene expression data normalization

Gene expression data was then normalized using Quantile normalisation (Matlab bioinformatics toolkit). This step made expression data in all databases directly comparable.

Differential expression analysis

Mean expression values and standard deviations were calculated for each set of tissues (high or low proliferation) or developmental windows with high or low proliferative activity during nervous system development. Comparisons of the collective level of expression of mitosis genes between conditions of high or low proliferative activity were carried out using both parametric (t-test) and non-parametic (Wilcoxon test) tests using the corresponding statistical functions in Matlab.

Co-expression and network structure analysis

Co-expression was calculated using Pearson correlation in Matlab using the 'corr' function.

This produced a matrix of correlation values. Using the function 'squareform' for this matrix the mean and standard deviation for each matrix was extracted. Pearson was used in all correlations; however Spearman was tested leading to essentially similar results. Similarly, mean was used in all average calculations and with the median (non-parametric tests)being tested as well, yielding essentially similar results.

Using the co-expression matrices in the different conditions, a threshold of 0.75 was chosen and used to convert the matrices to binary form. These binary matrices were imported into the Gephi software for network structure analysis (Bastian M. 2009).

The highly mitotic state was imported first, and the connections were allocated a value of 1. The low mitotic state was then appended onto the other plot, and given a value of 3. Connections that were present in both datasets were assigned the value of 2. Colours were assigned to connections by value; 1 = blue, 2 = yellow and 3 = red.

Differential Co-Expression Analysis

A script was constructed in R using the inbuilt WGCNA package (Langfelder and Horvath 2008).The genes belonging to each cluster were identified and analysed in isolation (see figures 5-8).

Results

In order to determine if average co-expression or expression is the most consistent signature of functional activation of mitosis associated genes we compared gene expression and co-expression during two differing states of mitosis activation. In the first instance we compared highly proliferative cell lines with nervous tissues. In a second analysis we compared two stages of nervous system development differing in their level of mitotic activity; highly proliferative developing nervous system vs less proliferative post-natal nervous system.

Average co-expression is a stronger signature of functional activation of mitosis-associated genes than average expression.

Gene expression data was obtained from BioGPS on tumoural cell lines (CL), normal non-nervous tissues (NNT) and nervous system tissues (NT) using microarray sampling (Table 1 and Table 2). Using gene ontology annotations, Mitosis associated genes and their corresponding expression data were isolated. Figure 2, panel A+B shows graphs representing change in mean expression and in panel C+D, co-expression (Purple Line ± S.E.M) of mitosis associated genes between the labelled pairs of tissue types (CL, NT or NNT). The orange lines in these charts represent the change in average expression of all other background genes. Co-expression was measured by obtaining the Pearson correlation from an equal number of samples (14) for each category of tissues.

These results show that both average expression and co-expression of mitosis-related genes are significantly higher in cell lines than in nervous tissues. However, when comparing non-nervous tissues with nervous tissues, we can see that only co-expression remains significantly different between these two

categories of tissues; suggesting that co-expression is a more consistent associated signature for mitosis-associated functional engagement.

Table 1 Dataset sources

Dataset	Source	Platform	Reference
1a	BioGPS – Human U133A/GNF1H Gene Atlas	RNA Microarray	[1a]
1b	BioGPS – Human NCI60 Cell Lines	RNA Microarray	[1b]
2	Gene Expression Omnibus (GEO) - NCBI	RNA Microarray	[2]
3	Brainspan	RNA-seq	[3]

Table 2: Tissues included in databases 1a and 1b according to proliferative activity

Non-Nervous Tissues (moderately proliferative)	Nervous Tissues (poorly proliferative)	Cell Lines (highly proliferative)
Appendix	Amygdala	CCRT CEM
CD14+ Monocytes	Caudate nucleus	DU145
CD19+ B Cells(Neg.Sel)	Cerebellum	HL60
CD33+ Myeloid	Cerebellum Peduncles	JURKAT
CD105+ Endothelial	Cingulate Cortex	K562
Kidney	Dorsal Root Ganglion	LNCAP
Liver	Globus Pallidus	MALME 3M
Lung	Hypothalamus	MOLT4
Placenta	Medulla Oblongata	RL7
Skeletal Muscle	Occipital Lobe	RPMI 8226
Skin	Olfactory Bulb	SKMEL2
Smooth Muscle	Parietal Lobe	SKMEL5
Testis	Pituitary	SKMEL28
Whole Blood	Spinal Cord	SR

Table 3: Tissues included in databases 2 and 3 (derived from microarray and RNA seq data respectively)

Tissue label	Brain Region
A1C	Primary Auditory Cortex
AMY	Amygdala
CBC	Cerebellar Cortex
DFC	Dorsolateral Prefrontal Cortex
HIP	Hippocampus
IPC	Posterior Inferior Parietal Cortex
ITC	Inferior Temporal Cortex
M1C	Primary Motor Cortex
MD	Mediodorsal Nucleus of the Thalamus
MFC	Medial Prefrontal Cortex
OFC	Orbital Prefrontal Cortex
S1C	Primary Somatosensory Cortex
STC	Superior Temporal Cortex
STR	Striatum
V1C	Primary Visual Cortex
VFC	Ventrolateral Prefrontal Cortex

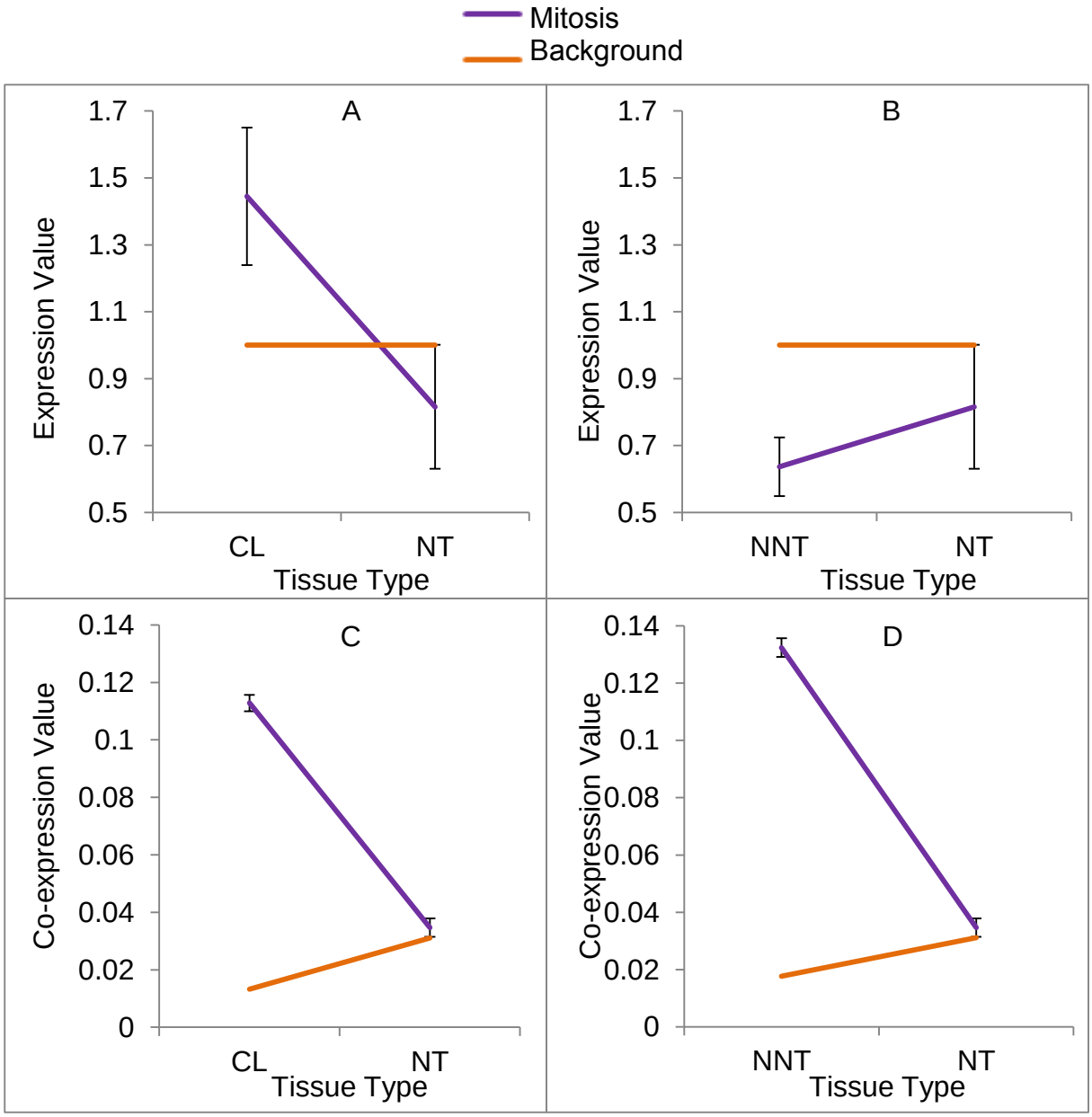


Figure 1: Average co-expression (examining cell lines, non-nervous tissues and nervous tissues) is a stronger signature of functional activation of mitosis associated genes than average expression.
Microarray gene expression data was obtained for tumoural cell lines (CL), normal non-nervous tissues (NNT) and nervous system tissues (NT). Using gene ontology annotations, Mitosis associated genes and their corresponding expression data were isolated. Panel A and B show graphs representing change in mean expression while panel C and D, co-expression (Purple Line ± SE) of mitosis associated genes comparing the indicated pairs of tissue types (CL, NT or NNT). The orange lines in these charts represent the change in average expression of background genes. Co-expression was measured by obtaining the Pearson correlation from an equal number of samples (14) for each category of tissues. (P Values for mitosis gene comparisons: A = $1.5x10^{-4}$, B = 0.117, C = $3.3x10^{-112}$, D = $1.2x10^{-79}$).

Mitosis associated genes display a stronger reduction in average co-expression than expression in the transition from pre-natal to post-natal brain development.

Gene expression data from four pre-natal samples (Pre) and post-natal samples (Post) were derived from the Gene Expression Omnibus database, generated using microarray profiling (Data set 2, table 1). Using gene ontology annotations, Mitosis associated genes and their corresponding expression data were isolated. Figure 2 panels A+B shows the change in mean expression (±S.E.M) and Figure 2 panels C+D shows the change in mean co-expression of mitosis associated genes between the different time frames (Purple Line). Figure 2; panel B and D, respectively, show the time course plot of average expression and co-expression (purple line) with the orange line representing the change in average expression or co-expression of background genes, respectively (4 months post conception through to 22 years). As before, co-expression was measured by obtaining the Pearson correlation from an equal number of samples and time points (16 brain regions x 8 time points = 128) for each of the above time frames. The time course of co-expression was measured as the Pearson correlation from an equal number of samples (16) per indicated time point in figures 2B and 2D.

The P value of the difference in expression when using a paired t-test comparison between pre and post-natal stages was $P = 9.4x10^{15}$ compared to the $P < 10^{-308}$ (the numerical lower limit of the statistical software used (Matlab)).

Our results show that average expression and co-expression of mitosis-associated genes are significantly higher in pre-natal development when compared with post-natal stages. Interestingly, when examining the time course of both expression and co-expression, we found that both expression and co-expression are consistently higher in pre-natal development before rapidly reducing after birth. In

either case the change in average co-expression of mitosis-associated genes in the perinatal transition was orders of magnitude more statistical significant than average level of expression.

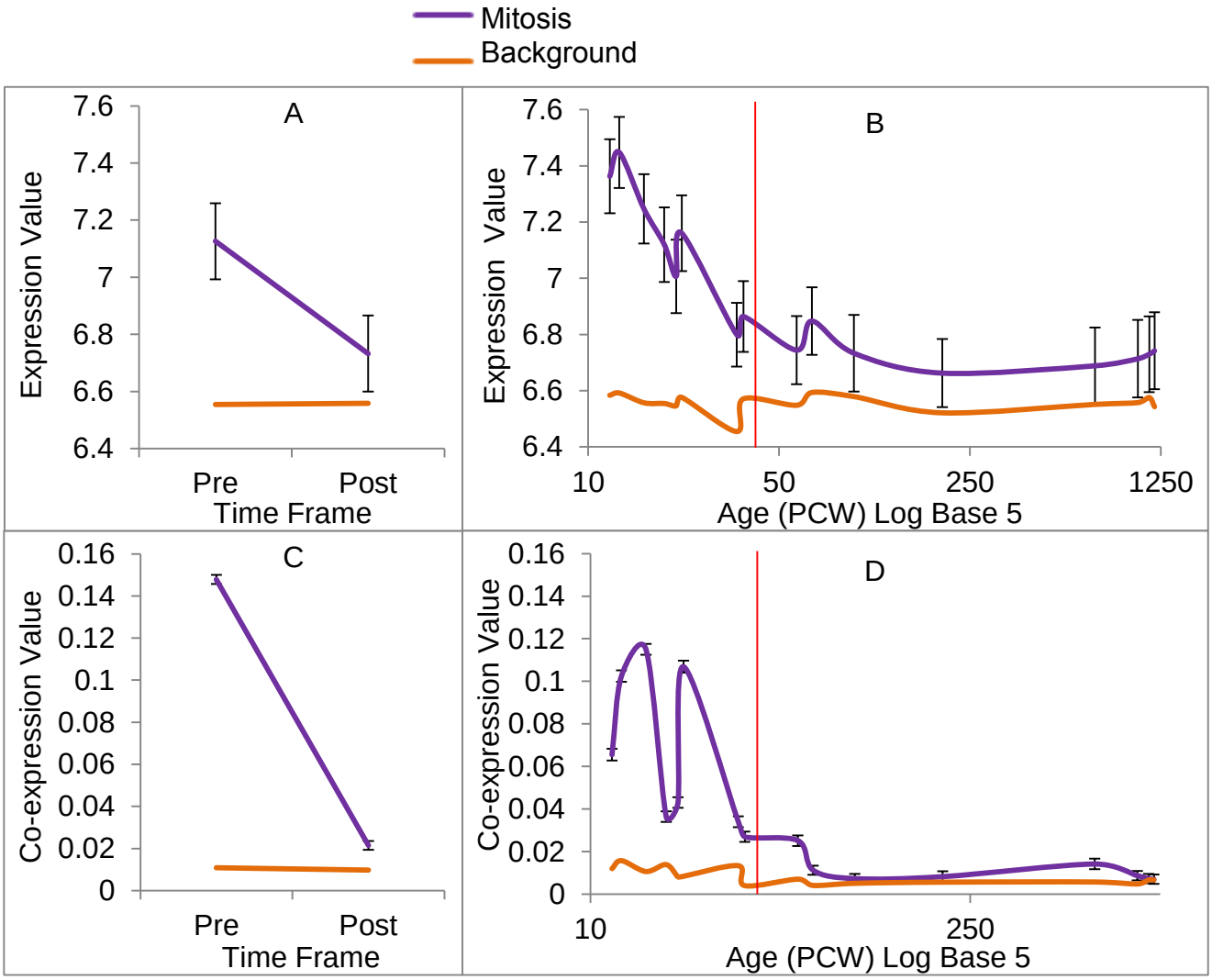
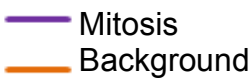


Figure 2: Mitosis associated genes display a stronger reduction in average co-expression than expression in the transition to post-natal brain development.
Microarray gene expression data from pre-natal samples (Pre) and post-natal samples (Post) were obtained from dataset 2 (see table 1). Using gene ontology annotations, Mitosis associated genes and their corresponding expression data were isolated.
A and B: Change in mean expression. C and D change in mean co-expression of mitosis associated genes between the different time frames (Purple Line) on the left, and the time course plot throughout development on the right. Orange line represents the change in average expression of all background genes. Co-expression was measured by obtaining the Pearson correlation from an equal number of samples and time points (16 samples x 8 time points = 128) for each of the above time frames. Time course co-expression was measured as the Pearson correlation from an equal number of samples (16) per time point. (Paired t-test p values for panels A and C: $9.4x10^{-15}$ and $<10^{-308}$ respectively).

These results demonstrate that intensity of co-ordinated expression of mitosis related genes is a much stronger signature of mitotic activity than their overall level of expression.

Mitosis-associated genes show a stronger reduction in co-expression than expression in transition to post-natal brain development when using RNA-seq data.

We confirmed the observed pattern of expression by examining independent, brain development expression based on next generation RNA-sequencing instead of microarray data. To this end we used the Brainspan dataset (dataset 4, table 1) which compiles RNA-sequencing data throughout human brain development from mid foetal to over 40 years of age. Gene expression data from three pre-natal stages (Pre: 16, 21 and 24 post-conception weeks) and three post-natal stages (Post: 19, 21 and 23 years) were obtained from this database. As before using gene ontology annotations, mitosis-associated genes and corresponding expression data were isolated. Figure 3, panel A and B shows the change in mean expression and panels C and D show the change in mean co-expression (±S.E.M) of mitosis associated genes between the different time frames as well as the time course of both mean expression and co-expression throughout development (16 PCW – 40 Years). Orange line represents the change in average expression of all other background genes. Co-expression was measured by obtaining the Pearson correlation from an equal number of samples and time points (16 brain region samples x 3 time points = 48) for each of the above time frames.



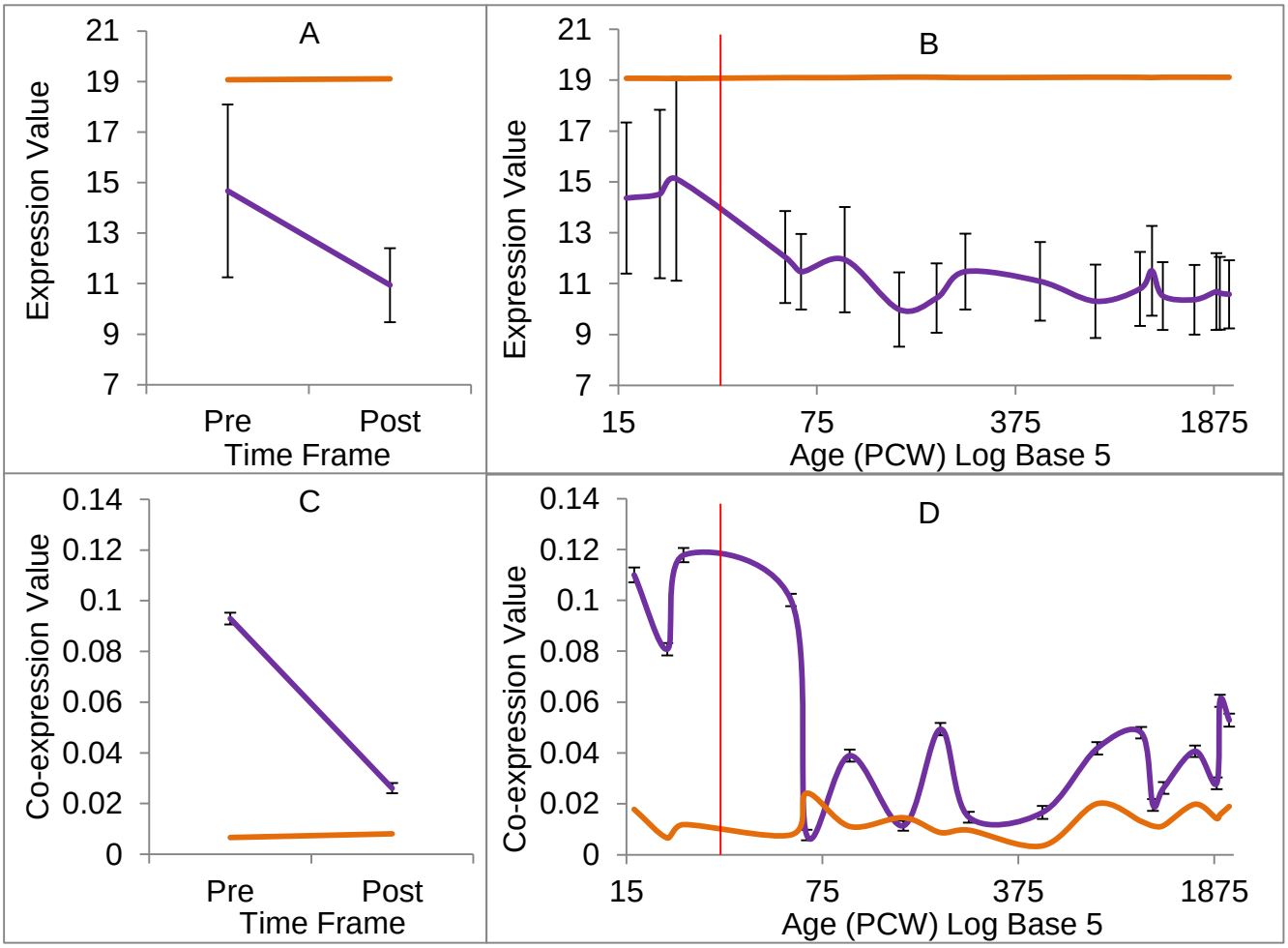


Figure 3: Line graphs showing that average co-expression (examining nervous tissues at different developmental stages) is a stronger signature, when compared to average expression, of functional activation of mitosis associated genes.

*Gene expression data from pre-natal samples (Pre) and post-natal samples (Post) were obtained from Brainspan.org database. Using gene ontology annotations, Mitosis associated genes and their corresponding expression data were isolated. [Top] Change in mean expression and [Bottom] change in mean co-expression of mitosis associated genes between the different time frames (Purple Line) on the left, and the time course plot on the right. Orange line represents the change in average expression of all background genes. Co-expression was measured by obtaining the Pearson correlation from an equal number of samples and time points (16 samples * 3 time points = 48) for each of the above time frames. Time course co-expression was measured as the Pearson correlation from an equal number of samples (16) per time point. (Paired t-test p values for panels A and C: 0.2076 and 1.710^{-130} respectively).*

As can be seen, the change in average expression level between pre-natal and post-natal stages was statistically non-significant (Paired t-test, P = 0.2076) whereas the corresponding change in average co-expression ($P = 1.7x10^{-130}$) was highly significant. These results confirm that co-expression of mitosis associated genes shows a much more pronounced change when comparing two developmental stages with differing demands for this function than average level of expression of the same set of genes.

Changes in the network structure of mitosis genes reflect differences in the proliferative demands of different tissues.

Changes in average co-expression could reflect either, homogenous variations in the degree or intensity of coordination across genes without changes in the regulatory structure of mitosis genes or, alternatively, these variations could result from an overall regulatory reorganization of mitosis genes in physiological conditions (or developmental stages) with differing demands for proliferative functions. In order to extract structural and regulatory information from the observed variations in co-expression, we conducted a detailed co-expression network analysis of mitosis genes comparing conditions of high and low demand of cell proliferation functions.

A co-expression network can be graphically represented as a set of nodes interconnected by edges or links; where nodes represent genes and links represent high co-expression between any two genes.

In order to determine the co-expression structure of mitosis-related genes we calculated the full correlation matrix for all possible pairs of mitosis-related genes for each of the previously described comparisons of physiological conditions,

namely: tumoural cell lines, non-nervous tissue, adult nervous tissue, pre-natal and post-natal brain expression. Any two genes were classified as connected whenever their correlation coefficient value was above 0.7 and comparisons of network structures were carried out between a) tumoural cell lines and nervous tissues, non-nervous vs nervous tissues; c) pre-natal vs post-natal brain using either microarray data or d) RNA-seq data.

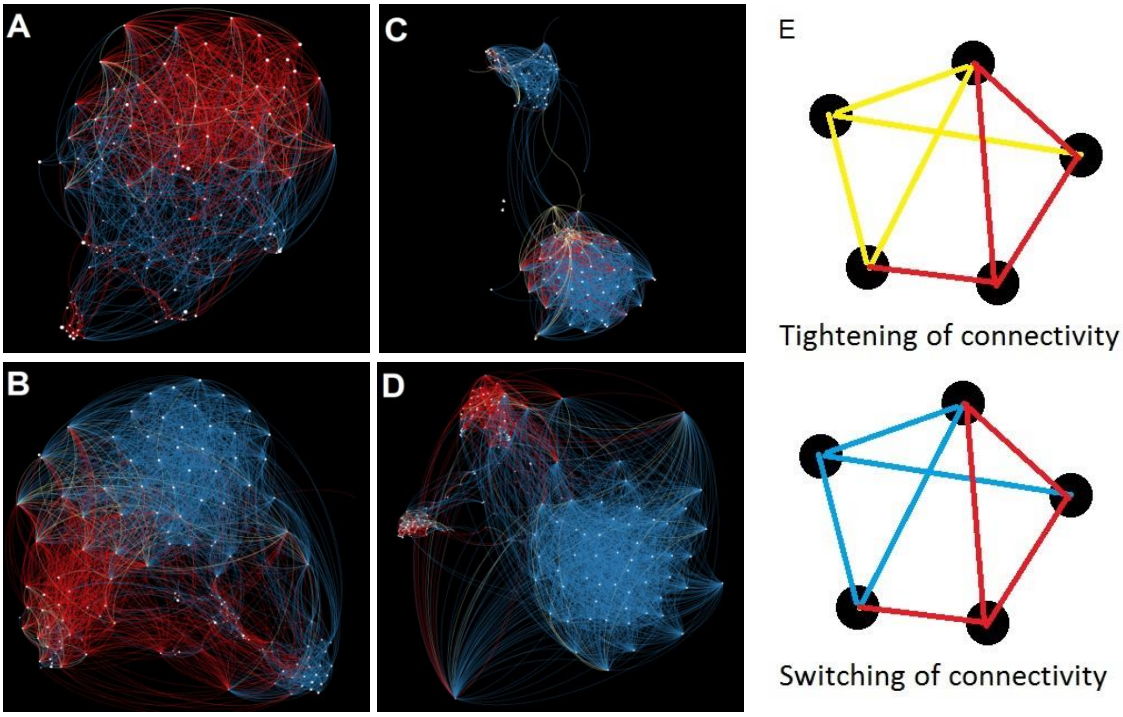


Figure 4: Structural changes of mitosis-associated gene co-expression networks comparing physiological conditions of high and low cell proliferation demands.

Connectivity of nodes based on co-expression value. Co-expression values of 0.7 or higher were classified as connected. Blue signifies high connectivity in high mitotic state, whereas red signifies high connectivity in the low mitotic state. Yellow represents constant high connectivity in both high and low mitotic conditions.
A: Structural changes of mitosis-associated gene co-expression networks comparing cell lines with nervous tissues. B: Structural changes of mitosis-associated gene co-expression networks comparing non-nervous tissues with less proliferative nervous tissues. C: Structural changes of mitosis-associated gene co-expression networks comparing highly proliferative pre-natal with lowly proliferative post-natal brain tissues (Microarray data). D: Structural changes of mitosis-associated gene co-expression networks comparing highly proliferative pre-natal with lowly proliferative post-natal brain tissues (RNA-seq data).
Node size is proportional to clustering coefficient value.
Panel E: Schematic representation of the expected network structure for a tightened or reorganized co-expression network structure.

Figure 4 shows graphic representations of these comparisons where blue links represent observed connections in the highly proliferative state, red links represent connections in the low proliferative state and yellow represent persistent connections in both high and low proliferative states.

An overall increase in the intensity of existing correlations, when comparing low to highly proliferative states, should result in the network showing mostly yellow links plus a few blue links corresponding to the new connection that would arise in the highly proliferative state. As can be seen in figure 4 all networks split into two definite domains or clusters of genes: those that interact only in the low proliferative state (joined by red links) and those that interact only in the highly proliferative state (blue links). These results show that the observed changes in correlated activity of mitosis associated genes are the result of an overall regulatory re-organization of mitosis genes in physiological conditions (or developmental stages) with varying demands for cell proliferation functions.

Differential co-expression analysis of mitosis-associated genes

The previous analysis shows that the observed differences in co-expression of mitosis genes in conditions of high and low functional demand are the result of an underlying reorganization of the regulatory network. In order to gain further insights into the detailed structural organization of these regulatory changes we used differential co-expression analysis as described by Tesson et al. (Tesson, Breitling et al. 2010). This method groups genes together when their correlations with the same sets of genes change between the different conditions. Briefly, we obtained the overall co-expression matrices for each of the comparative conditions described in the previous section and obtained the difference matrix resulting from subtracting the low proliferative condition from the highly proliferative condition. A

topological overlap matrix based on the differential co-expression matrix was then calculated follow by hierarchical clustering to identify modules of differentially co-expressed genes.

Differential co-expression analysis of pre and post-natal gene expression patterns in the human nervous system.

We started by conducting a differential co-expression analysis of the pre and post-natal gene expression patterns in the developing human brain. Hierarchical clustering of differentially co-expressed gene pairs led to the detection of eight distinct modules of gene clusters (Figure 5A). The corresponding comparative correlation heat map matrix of mitosis genes contrasting pre-natal and post-natal brain development reveals that most of the detected clusters display a high overall level of co-expression during pre-natal when compared to post-natal development (Figure 5B). In order to quantify this transition at the level of individual modules, we obtained the average co-expression of all pairs of genes within each module at both developmental stages. Figure 5C shows the average co-expression of each module (±S.E.M) and reveals that the previously observed overall developmental drop in mean correlation of all mitosis associated genes, is in fact the result of a regulatory reorganization occurring within separate and distinct gene clusters, with most of them reducing their level of correlated expression albeit at varying levels, with some even increasing their correlation (black module).

This result is consistent with Panel C from figure 4 as only a small number of connections are red. The red and pink clusters in figure 5 do not change co-expression significantly and the rest of the clusters decrease co-expression in the transition from pre-natal to post-natal development.

Differential co-expression analysis of pre and post-natal gene expression patterns in the human nervous system using RNA-seq data.

Using independent RNA-seq data set, we confirmed the observed pattern of modular re-organization in the regulatory structure of mitosis associated genes (Figure 6). Together these results suggest that observed overall changes in correlated activity of mitosis associated genes between pre-natal and post-natal brain development are the result of corresponding changes in correlated activity occurring within discreet gene clusters.

Differentil co-expression analysis comparing proliferative non-nervous tissues and less proliferative adult nervous tissues.

Next, we conducted a differential co-expression analysis comparing expression patterns of less proliferative adult nervous tissues with more proliferative non-nervous tissues. Hierarchical clustering of differentially co-expressed gene pairs led to the detection of seven distinct modules or gene clusters (Figure 7A). The corresponding comparative correlation heat map matrix of mitosis genes contrasting non-nervous and nervous gene expression patterns reveals again that most of the detected clusters display a high overall level of co-expression in the more highly proliferative non-nervous tissues than in the adult nervous tissue (Figure 7B). Figure 7C shows the average co-expression of each module (±S.E.M) and confirms a similar pattern to the one observed above where the overall change in mean correlation of all mitosis associated genes, was again the result of a regulatory reorganization occurring within separate and distinct gene clusters, with most of them displaying a higher level of correlated expression in highly proliferative non-nervous tissues compared to low proliferative nervous system tissues.

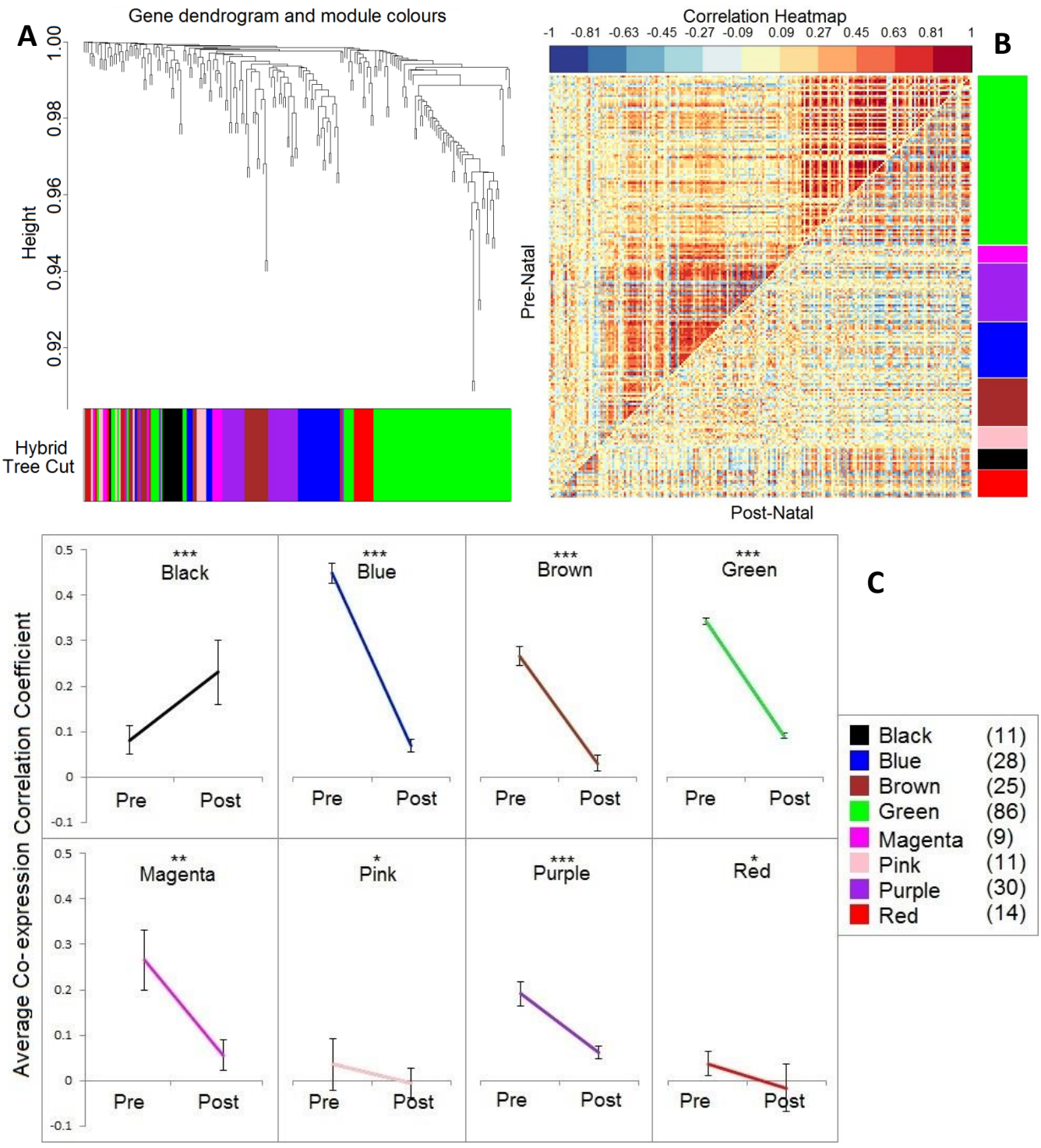


Figure 5: Differential co-expression analysis of pre and post-natal gene expression patterns in the human nervous system.

*A. Dendrogram tree diagram of differential co-expression clusters of mitosis gene pairs and associated module colours. B. Comparative correlation heat map matrix of mitosis genes comparing pre-natal and post-natal brain development. The upper diagonal shows pre-natal co-expression correlation values between pairs of mitosis genes while the lower diagonal shows co-expression correlations between the same pairs of genes at post-natal stages. Colour scaling indicated correlation coefficients from negative (blue) to positive (red) values. Modules are aligned with the module assignment colours on the right hand side. C. Charts showing changes in mean co-expression for each detected module. The value in brackets represents the number of genes within that cluster. Paired T-tests were used to assess statistical significance of the observed changes * P > 0.05, ** 0.05 => P > 0.005, *** P <= 0.005.*

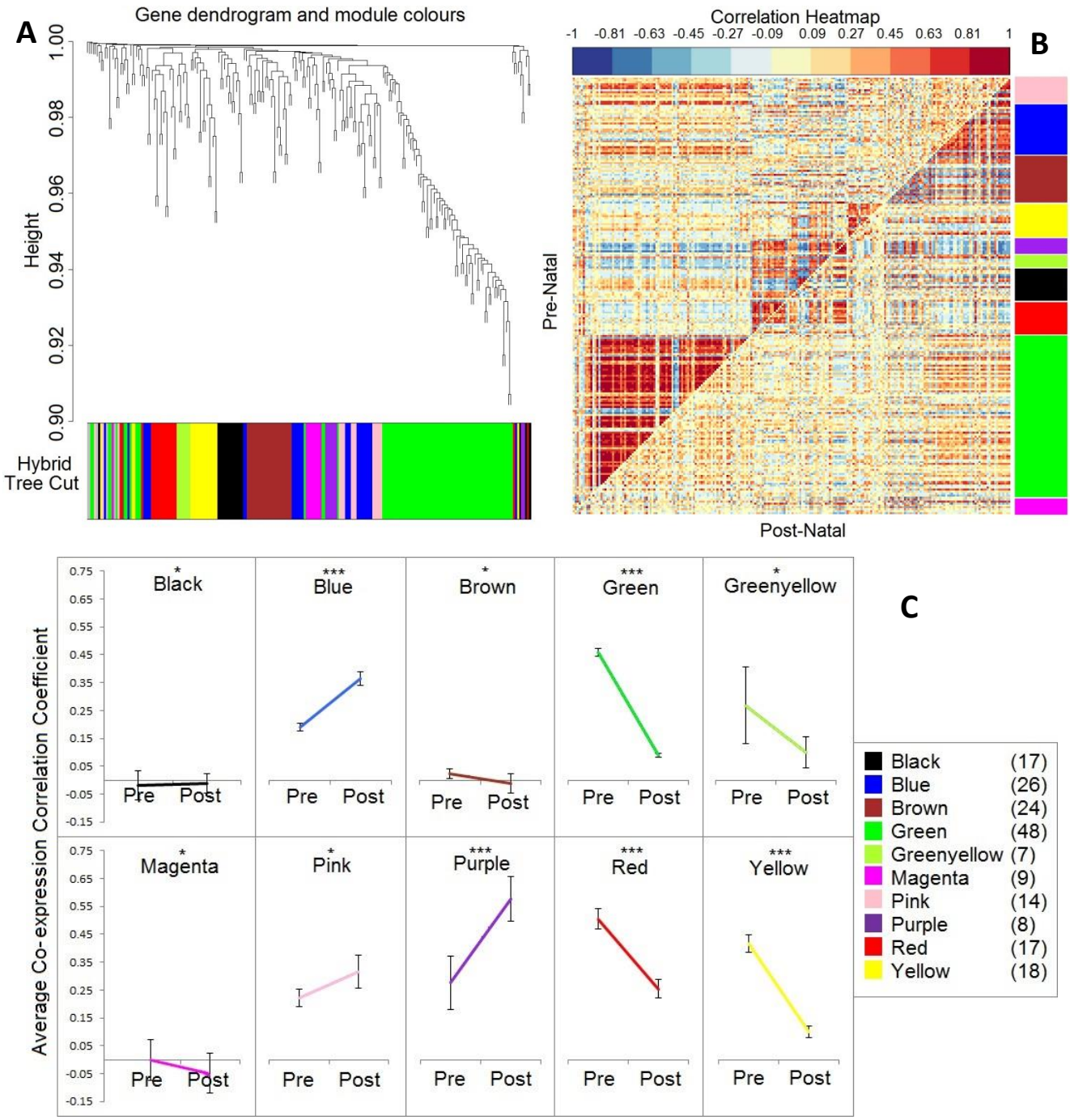


Figure 6: Differential co-expression analysis of pre and post-natal gene expression patterns in the human nervous system (RNA-seq data).

*A. Dendrogram tree diagram of differential co-expression clusters of mitosis gene pairs and associated module colours. B. Comparative correlation heat map matrix of mitosis genes comparing pre-natal and post-natal brain development. The upper diagonal shows pre-natal co-expression correlation values between pairs of mitosis genes while the lower diagonal shows co-expression correlations between the same pairs of genes at post-natal stages. Colour scaling indicated correlation coefficients from negative (blue) to positive (red) values. Modules are aligned with the module assignment colours on the right hand side. C. Charts showing changes in mean co-expression for each detected module. The value in brackets represents the number of genes within that cluster. Paired T-tests were used to assess statistical significance of the observed changes * P > 0.05, ** 0.05 => P > 0.005, *** P <= 0.005.*

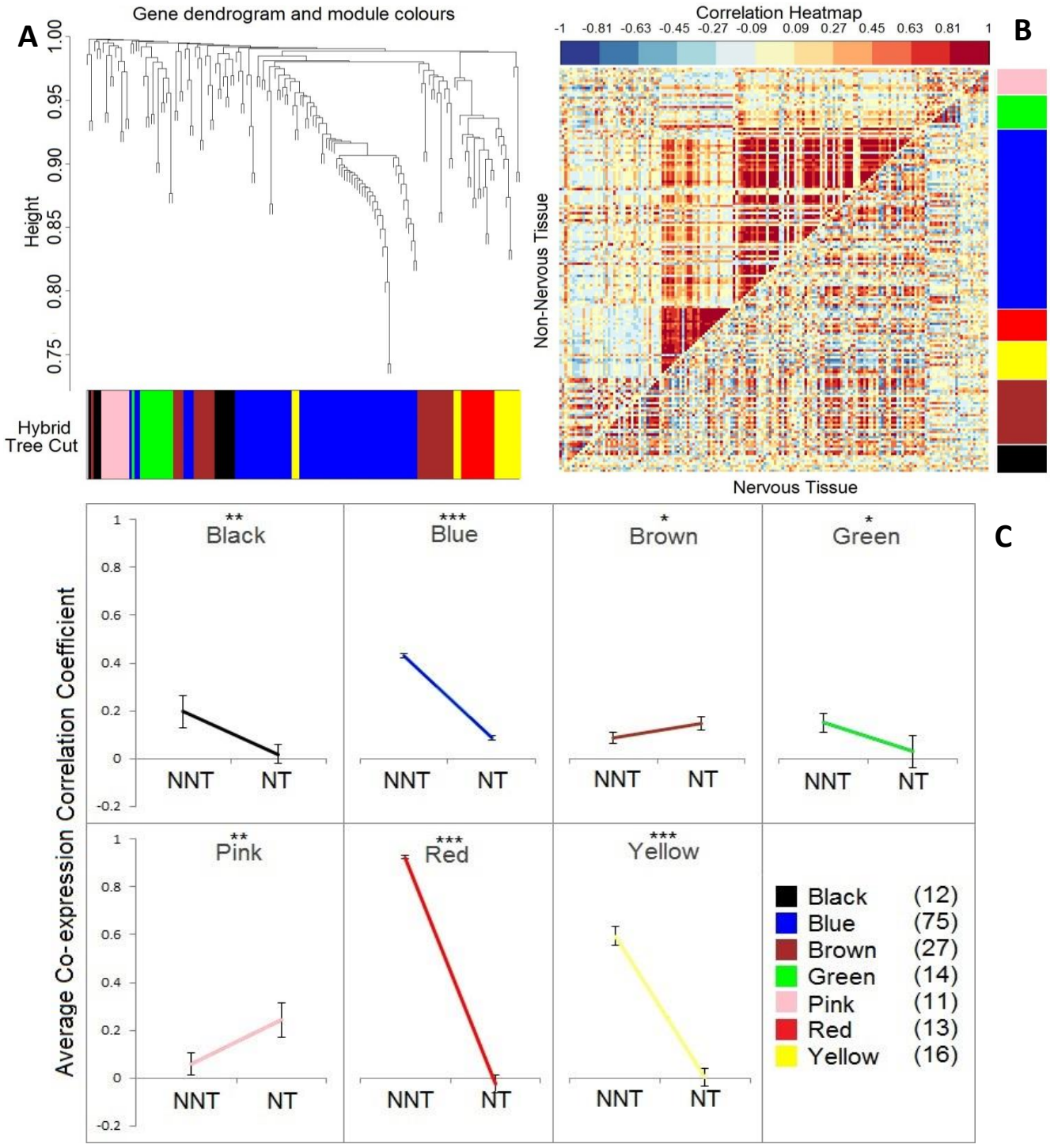


Figure 7: Differential co-expression analysis of non-nervous and adult nervous system gene expression patterns.

*A. Dendrogram tree diagram of differential co-expression clusters of mitosis gene pairs and associated module colours. B. Comparative correlation heat map matrix of mitosis genes comparing non-nervous and nervous tissues. The upper diagonal shows non-nervous co-expression correlation values between pairs of mitosis genes while the lower diagonal shows co-expression correlations between the same pairs of genes in nervous tissues. Colour scaling indicated correlation coefficients from negative (blue) to positive (red) values. Modules are aligned with the module assignment colours on the right hand side. C. Charts showing changes in mean co-expression for each detected module. The value in brackets represents the number of genes within that cluster. Paired T-tests were used to assess statistical significance of the observed changes * P > 0.05, ** 0.05 => P > 0.005, *** P <= 0.005.*

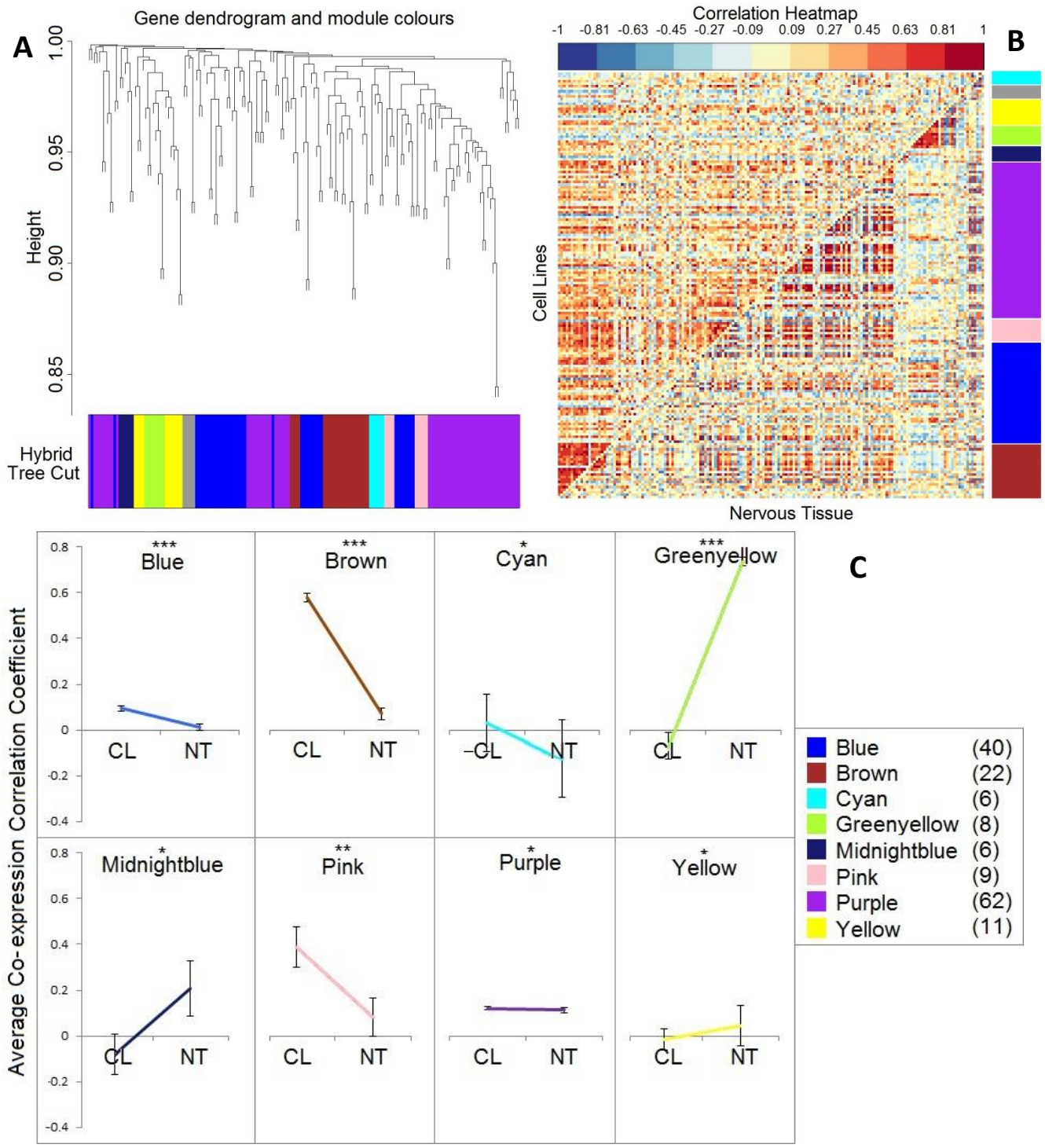


Figure 8: Differential co-expression analysis of comparing tumoural cell lines with adult nervous system gene expression patterns.

*A. Dendrogram tree diagram of differential co-expression clusters of mitosis gene pairs and associated module colours. B. Comparative correlation heat map matrix of mitosis genes comparing cell line and nervous tissues. The upper diagonal shows cell line co-expression correlation values between pairs of mitosis genes while the lower diagonal shows co-expression correlations between the same pairs of genes in nervous tissues. Colour scaling indicated correlation coefficients from negative (blue) to positive (red) values. Modules are aligned with the module assignment colours on the right hand side. C. Charts showing changes in mean co-expression for each detected module. The value in brackets represents the number of genes within that cluster. Paired T-tests were used to assess statistical significance of the observed changes * P > 0.05, ** 0.05 => P > 0.005, *** P <= 0.005.*

Differential co-expression analysis of gene expression patterns on cell line tissues and less proliferative adult nervous tissues.

A similar analysis comparing expression patterns of highly proliferative tumoural cell lines with less proliferative adult nervous tissues detected eight distinct modules or gene clusters (Figure 8A). The corresponding comparative correlation heat map reveals a more even distribution of clusters with five of them displaying more highly correlated expression in cell lines than in nervous tissues with the other three displaying the reverse trend (Figure 8C).

Discussion

Complex phenotypes at both organismal and cellular level are rarely the result of individual genes working in isolation. Most cellular functions are indeed the result of a large and complex assembly of molecular and genetic components acting in concert (Hartwell, Hopfield et al. 1999). Whether the engagement or activation of any cellular function is simply result of an overall up regulation or rather an increased regulatory coordination of the full assembly of associated genes is not known.

In this study, we looked at mitosis-associated genes and compared their level of expression and co-expression across cell types, physiological conditions and developmental stages known to differ in their level of proliferative activity. Mitosis-associated genes offer a uniquely simple opportunity to assess the relationship between overall pattern of expression and the actual engagement of proliferative functions. Cell proliferation is easy to assess and compare across different tissues, experimental conditions or developmental stages. To our knowledge, however, no previous study has addressed the link between the level of cell proliferation and the collective pattern of expression of the whole assembly of mitosis associated genes.

Specifically, we looked at expression data from a variety of human-derived tissues including human brain-derived samples at several developmental stages.

Using available microarray data, we compared highly proliferative cell lines with the much less proliferative nervous tissues and found that both average expression and co-expression of mitosis-associated genes were significantly higher in tumoural cell lines than in the nervous tissue samples in line with their lower demand of proliferative functions. Co-expression, however, was statistically

a much stronger indicator of functional engagement. The observed contrast is not the result of comparing the aberrant genetic behaviour of tumour-derived cells with normal tissues as normal highly proliferative non-nervous tissues such as endothelial tissue, kidney, liver, skin etc. showed significantly higher co-expression of mitosis genes than nervous tissues. Using both microarray and RNA sequencing expression data, we further compared average expression and co-expression of mitosis genes across pre-natal and post-natal brain-derived samples. In pre-natal development, cell proliferation demands are much higher than in post-natal stages as rapid and intense proliferation of neural precursors is required for early embryonic and late foetal formation of the brain. After birth this intense proliferative activity is considerably reduced as neural precursors virtually stop dividing and post mitotic neurons account for 50% of cells in the nervous system (Herculano-Houzel, Manger et al. 2014). Our results show that absolute level of expression did not significantly vary in a consistent way across these two developmental stages. This was particularly true when comparing microarray with RNA sequencing data. However co-expression of mitosis-associated genes was once again significantly higher during pre-natal development compared to post-natal stages, showing that co-expression is a much more robust signature of mitotic function requirement with average level of expression representing a less consistent predictor of functional activation.

The observed changes in average co-expression of mitosis-associated genes in line with increased demands of proliferative functions could reflect a corresponding increase in the degree of correlated activity across all possible pairs of mitosis-associated genes or, alternatively, the combination of increased and reduced regulatory interactions between mitosis genes in physiological conditions (or developmental stages) with differing demands for proliferative functions. While the

former would imply an overall tightening of all existing interactions, the latter would reveal rearrangements in the regulatory organization of mitosis-associated genes under varying demands of proliferative functions. In order to extract structural and regulatory information from the observed variations in co-expression, we conducted a detailed co-expression network analysis of mitosis genes comparing conditions of high and low demand of cell proliferation functions.

By examining the changes in the number of correlation interactions when comparing low to high proliferation demands we found that extremely few individual gene pairs remained highly correlated across conditions, thus ruling out a homogeneous tightening of all previously existing interactions. Instead, a modularity detection algorithm readily detects two discrete subpopulations of mitosis-associated genes: Those dissociating from each other as proliferative demands increase, and those (the majority) becoming increasingly associated in line with increased mitotic activity.

These results show that the observed changes in correlated activity of mitosis associated genes are the result of widespread rearrangements in the regulatory coordination of mitosis genes in physiological conditions of varying demands of cell proliferation functions.

We confirmed these findings by using a differential co-expression analysis (Tesson, Breitling et al. 2010) where distinct clusters of genes are specifically identified based on their similar pattern of differential co-expression relative to other genes. This analysis further reveals that mitosis genes undergo regulatory rearrangements in their pattern of expression of mitosis associated genes and that this rearrangement splits mitosis genes into discreet modules or clusters of differentially co-expressed genes, most of which increasing but some reducing their level of coordinated expression with increased demands of mitotic functions.

Taken together our results show that functional engagement of mitotic functions is much more strongly associated with increased regulatory coordination of mitosis-related genes than with their overall up regulation. Whether this association between functional engagement and co-expression can be extended to all biological functions is unclear, at least for now.

Circumstantial evidence, however, suggests that this could be the case. Indeed, groups of genes involved in a common function are likely to be under common regulatory control (Allocco, Kohane et al. 2004). Concerted expression of a given set of genes is normally brought about by the binding of a defined and rather limited set of transcription factors (Yu,Luscombe. 2003; Neph, Viestra.2012). In this regard it has been established that pairs of genes sharing common transcription factor binding sites, show on average higher co-expression than gene pairs that do not share common TF binding sites (Allocco, Kohane et al. 2004, Marco, Konikoff et al. 2009) suggesting that the joint activation of a set of genes involved in a common cellular function will be accompanied by an increased coordination in the pattern of expression of these genes and not necessarily their absolute level of expression.

This potentially general link between functional engagement and co-ordinated expression will certainly have wider implication for studies based on the simple detection of differential expression.

Genome wide transcriptome profiling studies tend to focus on the detection of differentially expressed genes across different conditions, cell types or developmental stages in order to identify the molecular basis of observed phenotypic differences (Gaitieri, Ding 2014). Our results however suggest that ongoing efforts to detect molecular assemblies' associated to specific cellular

phenotypes should focus in the increased co-expression of networks of genes rather than isolated events of up or down regulation of individual genes. Conversely, the level of activation of specific complex functions can be better characterized by the level of coordinated expression of a wider gene network known to be associated with those functions.

In summary our results show that co-expression is a much more robust signature of mitotic function than average level of expression and suggest that this association between functional engagement and co-expression can be extended to all biological functions.

References:

Allocco, D. J., I. S. Kohane and A. J. Butte (2004). "Quantifying the relationship between co-expression, co-regulation and gene function." BMC Bioinformatics **5**: 18.

Cooper, G. M. (2000). The Cell: A Molecular Approach. Sunderland (MA), Sinauer Associates.

Hartwell, L. H., J. J. Hopfield, S. Leibler and A. W. Murray (1999). "From molecular to modular cell biology." Nature **402**(6761 Suppl): C47-52.

Herculano-Houzel, S., P. R. Manger and J. H. Kaas (2014). "Brain scaling in mammalian evolution as a consequence of concerted and mosaic changes in numbers of neurons and average neuronal cell size." Front Neuroanat **8**: 77.

Gaiteri C, Ding Y, French B, Tseng GC, Sibille E: Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. Genes Brain Behav 2014, 13(1):13-24.

Langfelder, P. and S. Horvath (2008). "WGCNA: an R package for weighted correlation network analysis." BMC Bioinformatics **9**: 559.

Marco, A., C. Konikoff, T. L. Karr and S. Kumar (2009). "Relationship between gene co-expression and sharing of transcription factor binding sites in Drosophila melanogaster." Bioinformatics **25**(19): 2473-2477.

Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK et al: An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 2012, 489(7414):83-90.

Obayashi, T. and K. Kinoshita (2011). "COXPRESdb: a database to compare gene coexpression in seven model animals." Nucleic Acids Res **39**(Database issue): D1016-1022.

Oldham, M. C., S. Horvath and D. H. Geschwind (2006). "Conservation and evolution of gene coexpression networks in human and chimpanzee brains." Proc Natl Acad Sci U S A **103**(47): 17973-17978.

Oldham, M. C., G. Konopka, K. Iwamoto, P. Langfelder, T. Kato, S. Horvath and D. H. Geschwind (2008). "Functional organization of the transcriptome in human brain." Nat Neurosci **11**(11): 1271-1282.

Saris, C. G., S. Horvath, P. W. van Vught, M. A. van Es, H. M. Blauw, T. F. Fuller, P. Langfelder, J. DeYoung, J. H. Wokke, J. H. Veldink, L. H. van den Berg and R. A. Ophoff (2009). "Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients." BMC Genomics **10**: 405.

Tesson, B. M., R. Breitling and R. C. Jansen (2010). "DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules." BMC Bioinformatics **11**: 497.

Tolia, N. H. and L. Joshua-Tor (2006). "Strategies for protein coexpression in Escherichia coli." Nat Methods **3**(1): 55-64.

Torkamani, A., B. Dean, N. J. Schork and E. A. Thomas (2010). "Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia." Genome Res **20**(4): 403-412.

Tramontin, A. D., J. M. Garcia-Verdugo, D. A. Lim and A. Alvarez-Buylla (2003). "Postnatal development of radial glia and the ventricular zone (VZ): a continuum of the neural stem cell compartment." Cereb Cortex **13**(6): 580-587.

Usadel, B., T. Obayashi, M. Mutwil, F. M. Giorgi, G. W. Bassel, M. Tanimoto, A. Chow, D. Steinhauser, S. Persson and N. J. Provart (2009). "Co-expression tools for plant biology: opportunities for hypothesis generation and caveats." Plant Cell Environ **32**(12): 1633-1651.

Yu H, Luscombe NM, Qian J, Gerstein M: Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends Genet 2003, 19(8):422-427.

Zhang, B. and S. Horvath (2005). "A general framework for weighted gene co-expression network analysis." Stat Appl Genet Mol Biol **4**: Article17.

Databases:

Gene Expression Omnibus (GEO)-

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25219

BioGPS - http://biogps.org/downloads/

Brainspan - http://www.brainspan.org/static/download.html

Appendix

Mitosis genes and corresponding Ensembl gene IDs:

GO annotation	Ensembl ID
GO:0007067	ENSG00000003096
GO:0007067	ENSG00000007168
GO:0007067	ENSG00000008128
GO:0007067	ENSG00000010292
GO:0007067	ENSG00000011426
GO:0007067	ENSG00000011485
GO:0007067	ENSG00000019991
GO:0007067	ENSG00000037474
GO:0007067	ENSG00000047410
GO:0007067	ENSG00000053900
GO:0007067	ENSG00000058272
GO:0007067	ENSG00000062650
GO:0007067	ENSG00000064102
GO:0007067	ENSG00000066279
GO:0007067	ENSG00000068796
GO:0007067	ENSG00000068903
GO:0007067	ENSG00000070371
GO:0007067	ENSG00000072609
GO:0007067	ENSG00000072864
GO:0007067	ENSG00000101972
GO:0007067	ENSG00000102901
GO:0007067	ENSG00000103275
GO:0007067	ENSG00000104147
GO:0007067	ENSG00000105127
GO:0007067	ENSG00000105255
GO:0007067	ENSG00000105325
GO:0007067	ENSG00000107816
GO:0007067	ENSG00000108010
GO:0007067	ENSG00000108055
GO:0007067	ENSG00000108264
GO:0007067	ENSG00000110200
GO:0007067	ENSG00000110274
GO:0007067	ENSG00000111602
GO:0007067	ENSG00000111665
GO:0007067	ENSG00000111859
GO:0007067	ENSG00000112130
GO:0007067	ENSG00000113328
GO:0007067	ENSG00000113712
GO:0007067	ENSG00000113812

GO:0007067	ENSG00000075131
GO:0007067	ENSG00000075188
GO:0007067	ENSG00000079616
GO:0007067	ENSG00000080986
GO:0007067	ENSG00000084764
GO:0007067	ENSG00000086827
GO:0007067	ENSG00000087448
GO:0007067	ENSG00000087586
GO:0007067	ENSG00000088325
GO:0007067	ENSG00000089685
GO:0007067	ENSG00000090061
GO:0007067	ENSG00000090273
GO:0007067	ENSG00000091732
GO:0007067	ENSG00000092036
GO:0007067	ENSG00000094804
GO:0007067	ENSG00000094880
GO:0007067	ENSG00000100106
GO:0007067	ENSG00000100154
GO:0007067	ENSG00000100749
GO:0007067	ENSG00000101224
GO:0007067	ENSG00000101367
GO:0007067	ENSG00000101447
GO:0007067	ENSG00000101773

GO:0007067	ENSG00000114904
GO:0007067	ENSG00000115760
GO:0007067	ENSG00000116584
GO:0007067	ENSG00000116670
GO:0007067	ENSG00000117399
GO:0007067	ENSG00000117650
GO:0007067	ENSG00000117697
GO:0007067	ENSG00000117724
GO:0007067	ENSG00000118007
GO:0007067	ENSG00000119408
GO:0007067	ENSG00000119638
GO:0007067	ENSG00000119969
GO:0007067	ENSG00000120253
GO:0007067	ENSG00000120539
GO:0007067	ENSG00000121152
GO:0007067	ENSG00000121274
GO:0007067	ENSG00000121892
GO:0007067	ENSG00000122545
GO:0007067	ENSG00000122966
GO:0007067	ENSG00000123374
GO:0007067	ENSG00000124486
GO:0007067	ENSG00000127337
GO:0007067	ENSG00000127564

GO:0007067	ENSG00000128989
GO:0007067	ENSG00000129055
GO:0007067	ENSG00000129195
GO:0007067	ENSG00000129534
GO:0007067	ENSG00000129810
GO:0007067	ENSG00000130177
GO:0007067	ENSG00000130779
GO:0007067	ENSG00000131023
GO:0007067	ENSG00000131351
GO:0007067	ENSG00000132341
GO:0007067	ENSG00000133101
GO:0007067	ENSG00000133739
GO:0007067	ENSG00000135913
GO:0007067	ENSG00000136014
GO:0007067	ENSG00000136098
GO:0007067	ENSG00000136122
GO:0007067	ENSG00000136169
GO:0007067	ENSG00000136810
GO:0007067	ENSG00000137100
GO:0007067	ENSG00000137601
GO:0007067	ENSG00000137812
GO:0007067	ENSG00000137814
GO:0007067	ENSG00000138160
GO:0007067	ENSG00000138180

GO:0007067	ENSG00000145919
GO:0007067	ENSG00000146425
GO:0007067	ENSG00000146670
GO:0007067	ENSG00000146918
GO:0007067	ENSG00000147400
GO:0007067	ENSG00000147601
GO:0007067	ENSG00000147874
GO:0007067	ENSG00000149503
GO:0007067	ENSG00000149636
GO:0007067	ENSG00000149823
GO:0007067	ENSG00000149948
GO:0007067	ENSG00000150457
GO:0007067	ENSG00000152240
GO:0007067	ENSG00000152253
GO:0007067	ENSG00000153107
GO:0007067	ENSG00000153140
GO:0007067	ENSG00000154839
GO:0007067	ENSG00000156256
GO:0007067	ENSG00000156831
GO:0007067	ENSG00000156970
GO:0007067	ENSG00000157456
GO:0007067	ENSG00000158402
GO:0007067	ENSG00000159055
GO:0007067	ENSG00000159069

GO:0007067	ENSG00000138182
GO:0007067	ENSG00000138764
GO:0007067	ENSG00000139350
GO:0007067	ENSG00000140830
GO:0007067	ENSG00000140854
GO:0007067	ENSG00000141200
GO:0007067	ENSG00000141367
GO:0007067	ENSG00000141552
GO:0007067	ENSG00000141759
GO:0007067	ENSG00000142856
GO:0007067	ENSG00000142945
GO:0007067	ENSG00000143228
GO:0007067	ENSG00000143420
GO:0007067	ENSG00000143543
GO:0007067	ENSG00000143862
GO:0007067	ENSG00000143924
GO:0007067	ENSG00000144535
GO:0007067	ENSG00000144635
GO:0007067	ENSG00000145241
GO:0007067	ENSG00000145386

GO:0007067	ENSG00000160783
GO:0007067	ENSG00000161888
GO:0007067	ENSG00000162063
GO:0007067	ENSG00000162413
GO:0007067	ENSG00000163539
GO:0007067	ENSG00000163808
GO:0007067	ENSG00000163939
GO:0007067	ENSG00000164045
GO:0007067	ENSG00000164114
GO:0007067	ENSG00000164162
GO:0007067	ENSG00000164611
GO:0007067	ENSG00000164754
GO:0007067	ENSG00000165169
GO:0007067	ENSG00000165416
GO:0007067	ENSG00000165480
GO:0007067	ENSG00000166197
GO:0007067	ENSG00000166295
GO:0007067	ENSG00000166483
GO:0007067	ENSG00000166582
GO:0007067	ENSG00000166851

GO:0007067	ENSG00000166974
GO:0007067	ENSG00000167842
GO:0007067	ENSG00000167967
GO:0007067	ENSG00000168061
GO:0007067	ENSG00000168078
GO:0007067	ENSG00000168303
GO:0007067	ENSG00000168385
GO:0007067	ENSG00000169032
GO:0007067	ENSG00000169220
GO:0007067	ENSG00000169679
GO:0007067	ENSG00000169689
GO:0007067	ENSG00000170312
GO:0007067	ENSG00000171148
GO:0007067	ENSG00000173273
GO:0007067	ENSG00000173465
GO:0007067	ENSG00000174442
GO:0007067	ENSG00000174672
GO:0007067	ENSG00000175073
GO:0007067	ENSG00000175203
GO:0007067	ENSG00000175216
GO:0007067	ENSG00000175279
GO:0007067	ENSG00000175470
GO:0007067	ENSG00000175792
GO:0007067	ENSG00000176248

GO:0007067	ENSG00000198211
GO:0007067	ENSG00000167614
GO:0007067	ENSG00000198642
GO:0007067	ENSG00000198783
GO:0007067	ENSG00000198887
GO:0007067	ENSG00000198924
GO:0007067	ENSG00000203760
GO:0007067	ENSG00000204843
GO:0007067	ENSG00000213397
GO:0007067	ENSG00000214102
GO:0007067	ENSG00000214367
GO:0007067	ENSG00000249115
GO:0007067	ENSG00000250506
GO:0007067	ENSG00000258947
GO:0007067	ENSG00000262236
GO:0007067	ENSG00000262452
GO:0007067	ENSG00000262634
GO:0007067	ENSG00000263042
GO:0007067	ENSG00000263109
GO:0007067	ENSG00000268136
GO:0007067	ENSG00000268329
GO:0007067	ENSG00000268351
GO:0007067	ENSG00000268512
GO:0007067	ENSG00000268671

GO:0007067	ENSG00000176386
GO:0007067	ENSG00000176915
GO:0007067	ENSG00000177143
GO:0007067	ENSG00000179051
GO:0007067	ENSG00000180198
GO:0007067	ENSG00000182628
GO:0007067	ENSG00000182923
GO:0007067	ENSG00000183580
GO:0007067	ENSG00000183955
GO:0007067	ENSG00000184445
GO:0007067	ENSG00000184661
GO:0007067	ENSG00000185621
GO:0007067	ENSG00000185988
GO:0007067	ENSG00000186185
GO:0007067	ENSG00000186625
GO:0007067	ENSG00000186871
GO:0007067	ENSG00000196510
GO:0007067	ENSG00000197771
GO:0007067	ENSG00000198087
GO:0007067	ENSG00000269216
GO:0007067	ENSG00000269277

Used Scripts:

R:

DiffCoEx Tesson - http://pastebin.com/94wg7ubr

Matlab:

Co-Expression Matrix/Binary Conversion - http://pastebin.com/NEXEsZAj

Differential Co-expression – (Tesson, Breitling et al. 2010): R 3.0.2

```
1.  library(WGCNA)
2.  library(RColorBrewer)
3.  setwd("C:/WorkingDir")
4.  options(stringsAsFactors = FALSE);
5.
6.  #EXTRACT MODULES
7.  extractModules<-function(colorh1,datExpr,anno,write=F,file_prefix="",dir=NULL)
8.  {
9.    module<-list()
10.   if (!is.null(dir))
11.   {
12.     dir.create(dir)
13.     file_prefix=paste(dir,"/",file_prefix,sep="")
14.   }
15.   i<-1
16.   for (c in unique(colorh1))
17.   {
18.     module[[i]]<-(anno[colnames(datExpr)[which(colorh1==c)],1])
19.     if (write) {write.table(rownames(anno)[which(colorh1==c)],file=paste(file_prefix,"_",
    c,".txt",sep=""),quote=F,row.names=F,col.names=F)}
20.     i<-i+1
21.   }
22.   names(module)<-unique(colorh1)
23.   module
```

```
}


#PLOTC1C2 HEATMAP
plotC1C2Heatmap<-function(colorh1C1C2,AdjMat1C1,AdjMat1C2, X,
Y,ordering=NULL,file="DifferentialPlot.png")
{
  if (is.null(ordering))
  {
    h<-hclust(as.dist(1-
abs(cor(getEigenGeneValues(X[,which(colorh1C1C2!="grey")],colorh1C1C2[which(colorh1C1C2!=
"grey")],rbind(X,Y)[,which(colorh1C1C2!="grey")])))))
    for (c in h$label[h$order])
    {
      ordering<-c(ordering,which(colorh1C1C2 ==c))
    }
  }
  mat_tmp<-(AdjMat1C1[ordering,ordering])
  mat_tmp[which(row(mat_tmp)>col(mat_tmp))]<-
(AdjMat1C2[ordering,ordering][which(row(mat_tmp)>col(mat_tmp))])
  diag(mat_tmp)<-0
  mat_tmp<-sign(mat_tmp)*abs(mat_tmp)^(1/2)
  png(file=file,height=1000,width=1000)
  image(mat_tmp,col=rev(brewer.pal(11,"RdYlBu")),axes=F,asp=1,breaks=seq(-
1,1,length.out=12))
  dev.off()
  unique(colorh1C1C2[ordering])
}


#PLOTEXPRE CHANGE
plotExprChange<-function(X,Y, colorhC1C2,ordering=NULL)
{
```

```
53.   if (is.null(ordering))
54.   {
55.     h<-hclust(as.dist(1-
    abs(cor(getEigenGeneValues(X[,which(colorh1C1C2!="grey")],colorh1C1C2[which(colorh1C1C2!=
    "grey")],rbind(X,Y)[,which(colorh1C1C2!="grey")])))))
56.     for (c in h$label[h$order])
57.     {
58.       ordering<-c(ordering,which(colorh1C1C2 ==c))
59.     }
60.   }
61.   mycolors<-colorh1C1C2[ordering]
62.   plot(x=0:length(which(mycolors!="grey")),y=rep(1,length(which(mycolors!="grey"))+1),col
    ="white",axes=F,xlab="",ylab="",ylim=c(0,1))
63.   rr=c(244,239,225,215,209,193,181,166,151,130,110)
64.   gg=c(228,204,174,160,146,117,94,58,44,45,45)
65.   bb=c(176,140,109,105,102,91,84,74,70,68,66)
66.   MyColours<-NULL
67.   for ( i in 1:11)
68.   {
69.     MyColours=c(MyColours,rgb(rr[i],gg[i],bb[i],maxColorValue=255)  )
70.   }
71.   exprDiff<-NULL
72.   l<-0
73.   for (c in setdiff(unique(mycolors),"grey"))
74.   {
75.     meanC1<-mean(t(X)[colnames(X)[which(colorh1C1C2 == c)],])
76.     meanC2<-mean(t(Y)[colnames(Y)[which(colorh1C1C2 == c)],])
77.     exprDiff<-rbind(exprDiff,c(meanC1,meanC2))
78.     r<-l+length(which(mycolors==c))
79.     rect(l,0.85,r,1,col=c,border=F)
80.     rect(l,0,r,.4,col=MyColours[floor(meanC2*2)-10],border="white",lwd=2)
81.     rect(l,0.4,r,.8,col=MyColours[floor(meanC1*2)-10],border="white",lwd=2)
82.     l<-r
83.   }
```

```
84.   exprDiff
85. }
86.
87.
88. #GET EIGENGENE VALUES
89. getEigenGeneValues<-function(datRef,colorh1,datAll)
90. {
91.   eigenGenesCoef<-list()
92.   i<-0
93.   for (c in unique(colorh1))
94.   {
95.     i<-i+1
96.     eigenGenesCoef[[i]]<-prcomp(scale(datRef[,which(colorh1 == c)]))$rotation[,1]
97.   }
98.   names(eigenGenesCoef)<-unique(colorh1)
99.   values<-NULL
100.          for( c in unique(colorh1))
101.          {
102.            v<-rbind(datAll)[,which(colorh1 == c)] %*%  eigenGenesCoef[[c]]
103.            values<-cbind(values,sign(mean(v))*v)
104.          }
105.          colnames(values)<-unique(colorh1)
106.          values
107.        }
108.
109.
110.        X = read.csv(as.matrix("Prenatal Mitosis.csv",row.names=1, header=T,
    check.names=FALSE))
111.        Y = read.csv(as.matrix("Postnatal Mitosis.csv", row.names=1, header=T,
    check.names=FALSE))
112.        rownames(X)=X[,1]
113.        rownames(Y)=Y[,1]
114.        anno = as.character(X[,1])
115.        anno = as.matrix(anno)
```

```
116.        rownames(anno) = anno[1:215,]
117.        X = X[,2:129]
118.        Y = Y[,2:129]
119.        datC1 = t(X)
120.        datC2 = t(Y)
121.
122.
123.        beta1=6 #user defined parameter for soft thresholding
124.        MatC1 = cor(datC1,method="pearson")
125.        MatC2 = cor(datC2,method="pearson")
126.        AdjMatC1<-sign(MatC1)*(MatC1)^2
127.        AdjMatC2<-sign(MatC2)*(MatC2)^2
128.        AdjMat = (sqrt(.5*(abs((AdjMatC1)-(AdjMatC2)))))^beta1
129.
130.
131.
132.        TOM = TOMsimilarity(AdjMat)
133.        dissTOMC1C2 = 1 - TOM
134.
135.        geneTreeC1C2 = flashClust(as.dist(dissTOMC1C2), method = "average");
136.
137.        png(file="hierarchicalTree.png",height=1000,width=1000)
138.        plot(geneTreeC1C2, xlab="", sub="", main = "Gene clustering on TOM-based
    dissimilarity",labels = FALSE, hang = 0.04);
139.        dev.off()
140.
141.
142.        #We now extract modules from the hierarchical tree. This is done using
    cutreeDynamic. Please refer to WGCNA package documentation for details
143.        dynamicModsHybridC1C2 = cutreeDynamic(dendro = geneTreeC1C2,
    distM = dissTOMC1C2,method="hybrid",deepSplit = T,
    pamRespectsDendro =FALSE,minClusterSize = 5);
144.        #Every module is assigned a colour. Note that GREY is reserved for genes which do
    not belong to any differentially coexpressed module
```

```
145.        dynamicColorsHybridC1C2 = labels2colors(dynamicModsHybridC1C2)
146.
147.        #the next step merges clusters which are close (see WGCNA package documentation)
148.        mergedColorC1C2<-
    mergeCloseModules(rbind(datC1,datC2),dynamicColorsHybridC1C2,cutHeight=.1)$color
149.        colorh1C1C2<-mergedColorC1C2
150.
151.        # Plot the dendrogram and colors underneath
152.        png(file="module_assignment.png",width=1000,height=1000)
153.        plotDendroAndColors(geneTreeC1C2, colorh1C1C2, "Hybrid Tree
    Cut",dendroLabels = FALSE, hang = 0.03,addGuide = TRUE, guideHang = 0.05,main = "Gene
    dendrogram and module colors cells")
154.        dev.off()
155.
156.        #We write each module to an individual file containing affymetrix probeset IDs
157.        modulesC1C2Merged<-
    extractModules(colorh1C1C2,dissTOMC1C2,anno,dir="modules",file_prefix=paste("Output","Spe
    cific_module",sep=''),write=T)
158.        write.table(colorh1C1C2,file="module_assignment.txt",row.names=F,col.names=F,quot
    e=F)
159.
160.        #We plot to a file the comparative heatmap showing correlation changes in the
    modules
161.        #The code for the function plotC1C2Heatmap and others can be found below under
    the Supporting Functions section
162.
163.        plotC1C2Heatmap(colorh1C1C2,AdjMatC1,AdjMatC2, datC1, datC2)
164.        png(file="exprChange.png",height=500,width=500)
165.        plotExprChange(datC1,datC2,colorh1C1C2)
166.        dev.off()
167.
168.        rm(list=ls(all=TRUE))
```

```
%Enter a variable "Name" as the file name(without extension, assumed .csv)
X = csvread([Name,'.csv'],1,1);
Y = X';
Z = corr(Y);
Z(logical(eye(size(Z)))) = 0;
SQ = squareform(Z);
Mean = nanmean(SQ);
Std = nanstd(SQ);

csvwrite([Name,' MEAN.csv'],[Mean,Std])
csvwrite(['Matrix/',Name,' Matrix.csv'],Z)

%binary conversion
A = 1;
TH = 0.7; %threshold
THx = num2str(TH);
Genes = length(Z);
loop = 0;
while(loop<Genes^2)
    if(Z(A)>=TH)
        Z(A) = 1;
    elseif(Z(A)<TH)
        Z(A) = 0;
    end
    A = A+1;
    loop = loop+1;
end


csvwrite(['Matrix/Binary/',Name,' ',THx,' Binary.csv'],Z)
```